

Decomposing a Graph into Shortest Paths with Bounded Eccentricity

Etienne Birmelé¹, Fabien de Montgolfier², Léo Planche^{1,2}, and
Laurent Viennot^{3,2}

¹ MAP5, UMR CNRS 8145, Univ. Sorbonne Paris Cité

² IRIF, UMR CNRS 8243, Univ. Sorbonne Paris Cité

³ Inria

Abstract

We introduce the problem of hub-laminar decomposition which generalizes the one of computing a shortest path with minimum eccentricity (MESP). Intuitively, it consists in decomposing a graph into several paths that collectively have small eccentricity and meet only near their extremities. The problem is related to computing an isometric cycle with minimum eccentricity (MEIC). It is also linked to DNA reconstitution in the context of metagenomics in biology. We show that a graph having such a decomposition with long enough paths can be decomposed in polynomial time with approximated guaranties on the parameters of the decomposition. Moreover, such a decomposition with few paths allows to compute a compact representation of distances with additive distortion. We also show that having an isometric cycle with small eccentricity is related to the possibility of embedding the graph in a cycle with low distortion.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases Graph Decomposition, Graph Clustering, Distance Labeling, BFS, MESP

1 Introduction

The goal of this paper is to extend the MESP (Minimum Eccentricity Shortest Path) Problem from Dragan and Leitert [5] and the related problem of recognizing k -laminar graphs from Völkel *et al.* [16]. Both consist in finding a shortest path (in the sense that no path joining the same endpoints is shorter) k -dominating a graph (every vertex is at distance at most k from that path). The k -laminar problem additionally requires that path to be a diameter (there is no longer shortest path in the graph). Relationships between the two parameters are derived in [4].

To generalize this problem to more complex underlying structures, we introduce the problem of decomposing a graph into subgraphs with bounded shortest-path eccentricity. More precisely, we introduce the hub-laminar decomposition as a set of paths that k -dominates the graph and meet only near their extremities. To formalize this property, we introduce the notion of hub, that is a ball with fixed radius r centered at a path endpoint. The laminar associated to a path is the set of nodes k -dominated by the path. Our definition requires that an edge between two nodes belonging to two different laminars must also belong to a hub. The degree of a hub is then the number of laminars that meet in the hub. The main result of the paper is that computing such a decomposition becomes tractable when hub centers are far enough one from another, or equivalently when paths are long enough. The MESP problem is equivalent to a hub-laminar decomposition with one laminar.

Such a generalization is naturally interesting in networks where one might want to identify a set of speedy linear routes that are “highly accessible” with applications in communication networks, transportation planning and water resource management. It is also motivated by



© Etienne Birmelé and Fabien de Montgolfier and Léo Planche and Laurent Viennot;
licensed under Creative Commons License CC-BY

28th International Symposium on Algorithms and Computation (ISAAC 2017).

Editors: Yoshio Okamoto and Takeshi Tokuyama; Article No. 78; pp. 78:1–78:28

Leibniz International Proceedings in Informatics



LIPIC Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

DNA assembly in biology. DNA sequencing proceed through the reading of DNA fragments that must be assembled. When a single DNA strand is sequenced, comparison of fragments leads to a graph with “laminar” structure [16] that is with large diameter and small shortest path eccentricity. In the context of metagenomics, several DNA strands are sequenced together and more complex structures appear (see Figure 1 in [16]). Identifying the laminar structures of such graphs is typically encountered in metagenomic approaches for evolution questions (see e.g. [13]). The problem of the assembly (gluing DNA fragments to reconstruct a DNA strand) is then mixed with that of binning (sort DNA strands into groups that represent an individual genome or genomes from closely related organisms). See [14] for a presentation of assembly and binning problems in the context of metagenomics. Efficient decomposition of a graph into laminars could thus enhance the techniques for assembly and binning in this context.

The problem of decomposing a graph into λ laminars that k -cover the graph is not well defined as there may be several trade-offs of parameters λ and k . However, we show that when laminars are long enough compared to parameters r and k , then all (r, k) -hub-laminar decompositions are equivalent (same global structure) and have closely located hubs (except for hubs of degree two that do not affect the global structure). This implies for example that the positions of the extremities of the minimum eccentricity shortest path (MESP) can be approximated within $O(k)$ distance when the diameter of a graph is large with respect to the eccentricity k of the MESP.

From a graph perspective, a very natural generalization of MESP is the problem of finding a minimum eccentricity isometric cycle (MEIC), that is a cycle preserving distances that has minimum eccentricity k . Note that such a cycle can be seen as a hub-laminar decomposition with two laminars and two hubs with degree two. An important motivation for the MESP problem is its relationship with embedding a graph into the line with small multiplicative distortion [5]. We similarly show that the MEIC problem is related to embedding a graph into a circle with low multiplicative distortion, i.e. such that distances in the circle are within a constant factor of distances in the graph. Note that circle distortion is bounded by line distortion as a line segment can isometrically be embedded in a sufficiently long circle. (However, line distortion can be much larger than circle distortion.) Graph embedding in classical metrics is a well studied problem [9, 10]. Another related subject with abundant literature is that of compactly representing the distances of a graph [15, 12]. We show that a decomposition with few laminars ensures a compact representation of distances with bounded additive distortion.

Related works:

Finding a MESP is NP-complete but can be approximated within a constant factor [5]. Better trade-off between computation time and approximation factor for MESP is obtained in [4]. The problem of efficiently representing the distances in a graph encompasses a vast literature dating from metric embedding [1]. Approximating embedding with low distortion is introduced in [2] where some results are provided in the case of the line. The case of embedding the metric induced by an unweighted graph is studied in [3]. Embedding a graph metric into the line with minimum distortion is NP-complete but fixed parameter tractable with respect to distortion [6]. Approximate distance oracles, i.e. compact data-structures for representing an approximation of distances, are investigated in [15]. A particular approach introduced by Peleg [12] resides in assigning a label to each node of a graph such that the distance between two nodes can be estimated from their labels. Several result exist about the trade-off between label size and approximation quality. Exact distance estimation is

investigated in [8] and requires $\Omega(n)$ bits labels for general graphs. Approximation with a constant factor and sub-linear label size is derived in [15]. Some results concern additive approximation such as [7] in the case of hyperbolic graphs. A longest isometric cycle can be found in polynomial time [11].

2 Definitions

We consider finite, undirected and *connected* graphs (the connectivity is always assumed within the paper). Given a graph G , with vertex set $V(G)$ and edge set $E(G)$, we let $d_G(u, v)$ denote the *distance* between two vertices, i.e. the length of a shortest path from u to v . When the graph G is clear from the context, we omit the G subscript and simply write $d(u, v)$. Let $B(u, r) = \{v \in V(G) \mid d(u, v) \leq r\}$ denote the *ball* of radius r centered at u . Given a set of vertices U we set $B(U, r) = \cup_{u \in U} B(u, r)$. Given two sets U and W of vertices, we say that U *k-covers* W when every vertex in W is at distance at most k from some vertex in U , i.e. $W \subseteq B(U, k)$. We say that U has *eccentricity* k , denoted $\text{ecc}(U) = k$, when k is the smallest integer such that $B(U, k) = V(G)$. A path P in G is a sequence of nodes such that any two consecutive nodes are linked by an edge of G . We consider only simple paths: a node appears at most once in the sequence. The first node of the sequence and the last one are called the *endpoints* of P . For the simplicity of notations, we also let P denote the set of nodes appearing in the sequence. For any vertices u and v on P , we denote by P_{uv} the subpath of P having u and v as endpoints.

2.1 Hub-laminar decomposition

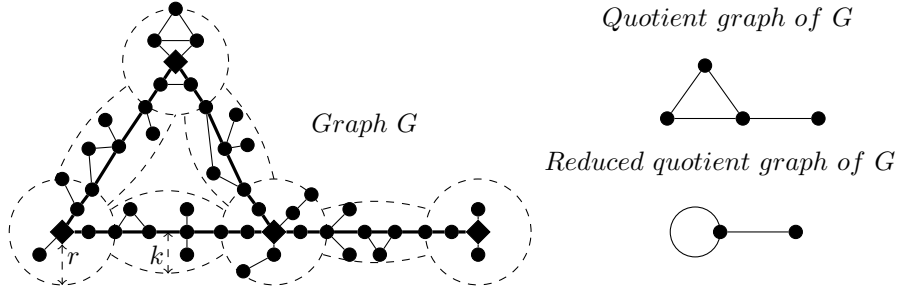
► **Definition 1** (Hub-laminar decomposition). Consider a connected undirected graph G , two positive integers r and k , $H = \{h_1, \dots, h_q\}$ a set of vertices of G called *hub centers*, and $\mathcal{P} = \{P_1, \dots, P_p\}$ a set of paths of G called *laminar paths*. A ball $B(h, r)$ with $h \in H$ is called a *hub*, and a set $B(P, k)$ with $P \in \mathcal{P}$ is called a *laminar*. (H, \mathcal{P}) is an (r, k) -*hub-laminar decomposition* of G if the following conditions are satisfied:

1. each laminar links two hubs centers: the endpoints h, h' of any $P \in \mathcal{P}$ belong to H and for every other hub $h'' \in H \setminus \{h, h'\}$, $B(P, k) \cap B(h'', r+1) = \emptyset$,
2. the laminars and the hubs cover G : $V(G) \subseteq \bigcup_{h \in H} B(h, r) \cup \bigcup_{P \in \mathcal{P}} B(P, k)$,
3. each laminar path is locally a shortest path: any path $P \in \mathcal{P}$ with endpoints h and h' is a shortest path of the graph $G[B(P, k) \cup B(h, r) \cup B(h', r)]$,
4. laminars meet at hubs only: for all $i \neq j$ and $uv \in E(G)$ such that $u \in B(P_i, k)$ and $v \in B(P_j, k)$, there is a hub center $h \in H$ such that P_i and P_j both have h as endpoint and $u, v \in B(h, r)$.

The *minimal laminar length* of a decomposition (H, \mathcal{P}) , denoted l , is the minimal length of the paths in \mathcal{P} . Its *laminar size*, denoted λ , is the number of paths in \mathcal{P} .

A hub-laminar decomposition (H, \mathcal{P}) with $l \geq 2r + 1$ forms a partition of the edges of G in the following sense: each edge is either inside exactly one hub (possibly touching many laminars ending in that hub), i.e. $\exists! h \in H$ s.t. $u, v \in B(h, r)$; or, else, inside a unique laminar (possibly touching one hub extremity of that laminar), i.e. $\exists! P \in \mathcal{P}$ s.t. $u, v \in B(P, k)$.

Figure 1 illustrates this definition and the notion of quotient graph that we define next. This definition basically defines a decomposition into k -neighborhoods of internally far apart shortest paths. It may seem a bit involved, but we think it expresses in a minimalist way what we mean by “internally far apart” with Item 4. Items 1 and 2 indicate that the graph is decomposed into laminars which are k -neighborhoods of certain paths and hubs which



■ **Figure 1** Illustration of an hub-laminar decomposition with $r = 2, k = 1$. Every vertex is at distance r from a hub center (diamond vertices) or at distance k from a laminar path (bold paths between hub centers).

are balls centered at the extremities of those paths. Item 3 requires path to be shortest in the induced graph, and not in G , to allow laminars with different length (otherwise, a long laminar between two hubs could be shortcut by one or more laminars forming a paths between these hubs).

2.2 Quotient graph and equivalence between decompositions

As previously mentioned, the hub-laminar decomposition gives naturally raise to a skeleton, which can be simplified into a *quotient graph*.

► **Definition 2** (quotient graph and reduced quotient). Given a graph G and an (r, k) -hub-laminar decomposition (H, \mathcal{P}) of G , the *quotient* of this decomposition is an edge-labeled multigraph with vertex-set H and for each $P \in \mathcal{P}$ with endpoints h, h' there is an edge hh' whose label is the length of P .

The *degree of a hub* denotes the degree of the corresponding vertex in the quotient graph, or equivalently the number of laminar paths its center is the endpoint of.

The *reduced quotient graph* of a decomposition (H, \mathcal{P}) is the multigraph obtained from its quotient graph by repeatedly removing degree 2 nodes: for every vertex u of the quotient incident with exactly two edges uv and uw with respective labels a and b , u and both edges are removed and a new edge vw is added with label $a + b$. (It is a loop when $v = w$.)

When the quotient is not a cycle (a case specifically addressed by MEIC, see Section 3) the reduced quotient is well defined and unique (recall graphs are supposed connected).

► **Definition 3** (equivalence between decompositions). Two hub-laminar decomposition of a same graph G , possibly with different parameters r, k , are *D-equivalent* if they have the same reduced quotient graph, up to an isomorphism ϕ of vertex-sets such that $d(h, \phi(h)) \leq D$ (d is the distance between hub centers in G , not in the reduced quotient)

2.3 Isometric cycle, circle embedding and distance labeling

A cycle C in a graph G is *isometric* if it preserves distances, i.e. $d_C(u, v) = d_G(u, v)$ for all $u, v \in V(C)$. In other words, for any pair u, v of nodes on the cycle, one of the two path linking u and v in the cycle is a shortest path in the graph. Note that an isometric cycle is necessarily an induced cycle. The MEIC problem consists in finding an isometric cycle with minimum eccentricity. It can be shown to be NP-complete following a similar proof as [5] for the NP-completeness of MESP problem.

A *circle embedding* of a graph G is a mapping $f : V(G) \rightarrow C$ where C is a circle of given length c . It has distortion γ if $d_G(u, v) \leq d_C(f(u), f(v)) \leq \gamma d_G(u, v)$ for all u, v in $V(G)$. The *circle distortion* $cd(G)$ of G is the minimum distortion of a circle embedding of G .

A distance labeling of a graph G consists in assigning a label L_u to each node $u \in V(G)$ together with a distance estimation function f that outputs an estimation of $d_G(u, v)$ when given L_u and L_v as input. It has additive distortion α if $d_G(u, v) \leq f(L_u, L_v) \leq d_G(u, v) + \alpha$ for all u, v in G .

3 Main results

Obviously, the reduced quotient graph of a graph having a (r, k) -hub-laminar decomposition follows the following trichotomy: it is either a path, a cycle or has a degree three node. We treat separately the three cases.

In the first case, the graph has a shortest path with eccentricity $\max\{3k, 2r\}$ and can be recognized through an approximate MESP algorithm such as [4]. (The $\max\{3k, 2r\}$ bound is a consequence of Lemma 12 given in Section 4.) In the second case, the graph has an isometric cycle with eccentricity at most $\max\{3k, 2r\}$. To recognize such graphs, we propose an approximate MEIC algorithm:

► **Theorem 4.** *Given a graph containing a K -dominating isometric cycle with length ℓ , a $6K$ -dominating isometric cycle can be computed in $O(n^{4.752} \log(n))$ time. Moreover, the computed cycle is indeed $3K$ -dominating when $\ell \geq 12K + 2$.*

We obtain therefore an algorithm for approximating circle embedding with low distortion.

► **Corollary 5.** *If a graph has circle distortion γ , it is possible to embed it in a circle with distortion $O(\gamma^2)$ in polynomial time.*

Recognizing the general case of decomposition is not a well defined problem as several decompositions may yield different trade-offs of the parameters. However, when laminars are long enough, all (r, k) -hub-laminar decompositions are indeed $O(k)$ equivalent. This can be seen as a consequence of the following recognition result.

► **Theorem 6.** *Given a graph G having a (r, k) -hub-laminar decomposition (H, \mathcal{P}) of minimal laminar length $\ell \geq 8r + 60k + 4$ and integers K, R such that $K \geq 3k$, $R \geq 2K + 3r + 3k$ and $2R + 8K < \ell - 2r - 18k - 4$, it is possible to compute in $O(\min(n, \lambda)m)$ time a (K, R) -hub-laminar decomposition which is $(K + 2r)$ -equivalent to (H, \mathcal{P}) .*

From the graph metric point of view, we obtain then a compact representation of distances:

► **Corollary 7.** *Given a graph G having an (r, k) -hub-laminar decomposition with laminar size λ , it is possible to compute in polynomial time a $O(\max\{k, r\})$ -additive distance labeling with $O(\lambda \log n)$ bit labels.*

Due to lack of space, the proofs of these theorems, and of the lemmas and propositions stated below, are put in Appendix.

4 Algorithms

4.1 Minimum Eccentricity Isometric Cycle

We propose to approximate the MEIC problem by computing a longest isometric cycle, that is an isometric cycle of G with maximum length. The following lemma shows that a longest isometric cycle $O(k)$ -dominates any k -dominating isometric cycle.

► **Lemma 8.** *Let G be a graph with an isometric cycle $C = c_1, \dots, c_p$ k -dominating G , and let D be a longest isometric cycle of G . Every vertex of C is at distance at most $5k$ of D . Furthermore, if D has length more than $12k + 2$ then every vertex of C is at distance at most $2k$ of D .*

Consequently, a longest isometric cycle in a graph is a 6-approximation for the MEIC problem, and a 3-approximation when the graph has a diameter large enough. As shown in [11], a longest isometric cycle can be computed in $\mathcal{O}(n^{4.752} \log(n))$ time. Theorem 4 is thus a direct consequence of this and Lemma 8.

4.2 General case outline

Consider a graph G having a (r, k) hub-laminar decomposition (H, \mathcal{P}) of minimal laminar length ℓ and having at least one hub of degree at least 3. The underlying idea of the algorithm is to use BFS (Breadth-first search) to compute shortest paths and their K -neighborhoods, K being chosen large enough to dominate every laminar traversed by the considered shortest paths, but small enough compared to ℓ to detect all hubs of degree at least 3.

The first step, called *FindHubs*, consists in applying the procedure *NextHub* described in section 4.4 until it discovers no new hubs. This step yields two sets of hub-centers A and B , respectively called *unmovable* and *movable* hub-centers, which will be used to determine the laminars. A unmovable hub center $a \in A$ means we are sure there exists exactly one hub center $h \in H$ such that $d(a, h)$ is bounded. It will be shown that A contains exactly one such vertex for every hub-center of H which degree is not 2.

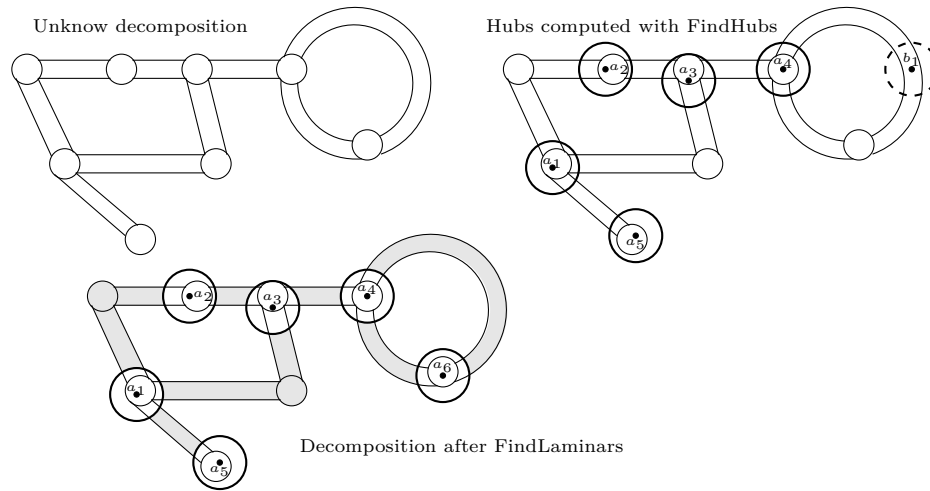
A movable hub center $b \in B$ will only be added by *NextHub* in a configuration corresponding to a cycle in the quotient graph of (H, \mathcal{P}) containing only one hub of degree at least 3, like the three laminars on the left of Figure 1. This is called a *Problematic Configuration*. We know there exists at least Degree 2 hub $h \in H$ somewhere in that cycle, but if they are thin enough they may remain merged in the laminars and we can not bound $d(b, h)$, and in the second step b may be moved.

The laminars are determined in a second step by the *FindLaminars* procedure, which links the hub-centers of the previous step by shortest paths. The only difficulty which has to be taken into account refers to hubs of degree 2 in (H, \mathcal{P}) . Indeed, the BFS runs of the hub-detection step may have missed one of them because they K -dominated it, whereas the BFSs of second step don't. In that case, the set of hubs A is adapted by adding the new discovered hub, and if needed, the corresponding movable hub center is deleted from B .

Figure gives a summary of the two steps by showing a possible outcome of the *FindHubs* and *FindLaminars* on an example. The *FindHubs* procedure detects all hubs of degree different from 2 and some of those of degree 2. Moreover, it places a movable hub on each problematic configuration. *FindLaminars* then computes the corresponding laminars, adding new hubs if a hub of degree 2 missed in the first step is detected. Some of them may however still be undetected, being replaced by a movable hub or just missing in the final decomposition. The quotient graphs of the decomposition supposed by Theorem 6 and of the constructed decompositions may therefore be different, but their reduced quotients are equal.

4.3 Some rules to compute a decomposition

We rely on the following properties for running our algorithm. The first tool is used to identify hub centers. Fortunately there is a pattern that, when it occurs, signals that any hub-laminar decomposition must have a hub nearby:



■ **Figure 2** Illustration of the different steps of the algorithm. The (H, \mathcal{P}) decomposition is unknown (top left). Notice a Problematic Case on the right of the graph: a cycle of laminar with only one degree not 2 hub. First, hub centers are computed such that every degree not 2 hub $B(h, r), h \in H$ is covered by $B(a_i, R), a_i \in A$ (top right). Finally the laminar are computed (greyed, bottom) and some movable hubs may be moved (like b_1 moved into a_5). Some thin degree 2 hubs from H are not found but merged in the K -laminars. Fortunately the hub center a_5 was found by BFS in the second step, but we could also have output b_1 instead, or both a_5 and b_1 , yielding in any case an equivalent reduced quotient.

► **Lemma 9** (Hub trigger). *Consider three numbers r, k and $K \geq 3k$. If there exists*

- *a shortest path Q from a to b*
 - *a vertex $u \in V(Q)$ such that $d(a, u) > K + 6k$ and $d(b, u) > K + 6k$*
 - *a vertex v such that $d(u, v) = K$*
 - *a vertex w such that $vw \in E(G)$ and $d(Q, w) = K + 1$ and $d(u, w) = K + 1$,*
- then any (r, k) -hub-laminar decomposition (H, \mathcal{P}) has a hub center $h \in H$ with $d(u, h) \leq K + r$.*

Fig. 3 (right) illustrates this. This pattern, when found, allows to propose u as an hub. And it is very powerful, since every hub h of degree at least three of any (r, k) -hub-laminar decomposition shall trigger this pattern, for any shortest path Q passing close to h and long enough, as the following lemma says:

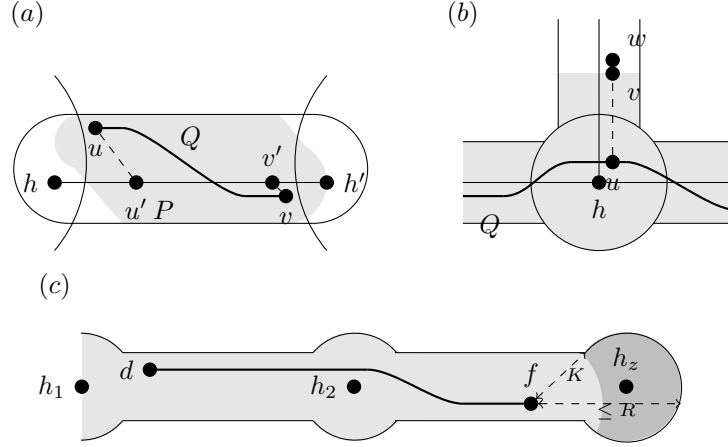
► **Lemma 10** (Degree ≥ 3 Hub Detection). *If a graph admits an (r, k) -hub-laminar decomposition (H, \mathcal{P}) , and has a hub $h \in H$ whose degree is at least 3, then for any*

- *$K \geq 3k$*
- *shortest path Q from a to b*
- *vertex $u \in V(Q)$ such that $d(a, u) > r + 4K + 9k + 2$ and $d(b, u) > r + 4K + 9k + 2$ and $d(u, h) \leq K$*

there exists

- *a vertex $x \in V(Q)$ such that $d(a, x) > K + 6k$ and $d(b, x) > K + 6k$*
- *a vertex v such that $d(x, v) = K$*
- *a vertex w such that $vw \in E(G)$ and $d(Q, w) = K + 1$ and $d(x, w) = K + 1$,*

Notice that if a graph admits an (r, k) -hub-laminar decomposition (H, \mathcal{P}) where all hubs have degree at least three, then the pattern is enough to find all its hubs, or more exactly to compute a set of hubs H' which is in bijection ϕ with H , i.e. for all $h \in H$ $d(h, \phi(h)) \leq K + r$.



■ **Figure 3** Illustration of the structural properties. (a) Lemma 12: The part of the laminar k -covered by $P_{u'v'}$ is K -covered by Q . (b) Lemma 9: If Q goes through a hub of degree ≥ 3 , a triplet u, v, w can be found, and it is the only such case far apart the extremities of Q . (c) Lemma 11.

Of course to do so in polynomial time we shall use a clever collection of paths Q to trigger all hubs. This is the idea developed in the algorithm. But before we shall explain how to deal with degree 1 hubs, the “dead end” hubs.

► **Lemma 11** (Hub in the dead-end). *Consider the graph G' induced by a sequence of incident hubs and laminar $H_1, L_1, H_2, \dots, H_z$, such that h_1 and h_z are at distance of at least $2R + r + 2$. Suppose moreover that H_z is a hub of degree 1 and all other hubs but H_1 are of degree 2.*

Let d in L_1 be at distance at most $R + r$ of h_1 and f a vertex of G' the furthest from d . f is then at distance at most $2r + 2k$ from h_z .

This lemma allow to approximate h with f . As we have just seen, hub of degree not 2 are, in some sense, uniquely defined (up to a certain distance) in any hub-laminar decomposition of given parameters. Degree 2 hubs however may be added at discretion on any hub-laminar decomposition, in the middle of long laminar, so we cannot imagine a sufficient condition for detecting them. However, they are necessary in very few case, namely

- to dominate a vertex at distance more than k , but less than r , inside of a r -laminar (not k -laminar)
- for the Problematic Configuration, since a laminar must have two distinct extremities

The last property, proved in [4], deals no more with computing hub centers but with computing laminars. While is NP-hard to find a shortest path that k -dominates a k -laminar graph, any path $3k$ -dominate a section of the laminar between its extremities. More formally

► **Lemma 12** (Path local dominating). *Consider a shortest path P (say, from h to h'). Let Q be a path from u to v contained in $B(P, k)$.*

Assume there exists $u' \in P$ and $v' \in P$ such that $d(u, u') \leq k$ and $d(v, v') \leq k$.

Then every vertex of $P_{u'v'}$ is at distance at most $2k$ from Q .

Furthermore, every vertex of $B(P_{u'v'}, k)$ is at distance at most $3k$ of Q .

Fig. 3 (left) illustrates this. We extensively use this lemma for designing an approximation algorithm: P is any laminar path, and Q is chosen to $3k$ dominate the middle of the laminar of P , i.e. all vertices far enough from P extremities (Lemma13 and 14 define “far enough” as $2R + 8K + 2r + 18k + 4$) and we therefore get a decomposition into $3k$ -laminar graphs.

4.4 Finding hubs

4.4.1 The StopBFS function

Let G be a vertex-colored graph with some uncolored vertices. The **StopBFS** procedure, provided a vertex d and a color c , consists in running an usual Breadth-first search, starting at vertex d , with the following additional rules:

- only vertices without color c are put in the BFS queue
- the BFS stops immediately if a vertex f is visited (i.e. extracted from BFS queue) and f has a colored neighbor whose color is not c .
- otherwise, if the BFS stops because its queue is empty, let f be the last visited vertex
- function $StopBFS(d, c)$ returns the BFS path P from d to f (which is a shortest path in the graph induced by G after removing c -colored vertices).

4.4.2 Finding a new hub: NextHub

Given a vertex s , typically corresponding to an already selected hub center, the **NextHub** procedure (see pseudo-code below) detects new hubs: it colors $B(s, R)$ with a next color and runs a StopBFS procedure from its border. In the case of a not deep-enough tree, the discovered vertices are colored to not be reused during the hub discovery. Otherwise, it may either find a new hub of degree at least 3 by Lemma 10, find a new hub of degree 1 by Lemma 11, meet another hub and dominate a lamina by Lemma 12 or cycle and come back to hit $B(s, R)$. The latter case indicates that the algorithm encountered the problematic configuration and induces the creation of a movable hub.

Given a path P , $r3K(P)$ denotes the subpath of P get by removing the $3K$ first and $3K$ last vertices of P . In the latter, sets A and B respectively denote the unmovable and movable hub centers.

4.4.3 Finding all hubs : FindHub

The *FindHub* simply consists in considering the initial uncolored graph G and to construct the sets A and B of unmovable and movable hub-centers by repeatedly applying *NextHub* (see pseudo-code in Appendix).

For the first call, we first compute a long path Q using a double BFS. More precisely, starting at any s_0 , we compute furthest node s and then repeatedly apply *NextHub* until a vertex a is added to A . If the hub of degree at least three is unique, the fact that $\ell > 2R + 8K + 2r + 18k + 4$ ensures that the deepest vertex of any BFS is at distance greater than $R + (r + 4K + 9k + 2)$ of the hub center. If there are at least two hubs of degree at least three, $\ell > 2R + 8K + 2r + 18k + 4$ implies that any vertex is at distance greater than $\frac{\ell}{2} > R + (r + 4K + 9k + 2)$ of one of the two hub-centers. In any case, the *Next_Hub* function applied to s has to find the configuration from Lemma 9 at some point, ensuring that a first hub center $a \in A$ is found. We set $A = \{a\}$ and $B = \emptyset$, and uncolor the whole graph.

Once this first vertex of A has been found, *NextHub* is run while there exists a hub center $a \in A$ having an uncolored vertex in its $R + 1$ -neighborhood. If ℓ is large, the *FindHubs* procedure finds all hubs of (H, \mathcal{P}) up to those of degree 2, as stated in the following lemma.

► **Lemma 13.** *Suppose that (H, \mathcal{P}) has at least a hub of degree 3, and $\ell(H, \mathcal{P}) > 2R + 8K + 2r + 18k + 4$. Then, for every vertex in $a \in A$, there exist a vertex in $h \in H$ such that their distance is at most $K + 2r$. Conversely, for every $h \in H$ of degree different from 2, such a vertex a is selected in A .*

```

1 NextHub
   Input: A graph  $G$  with possibly colored vertices, integers  $R$  and  $K$ , hub-center
           sets  $A$  and  $B$ , and a vertex  $s$ 
   Output: Updated sets  $A$ ,  $B$  and vertex coloring
2 Color every vertex in  $B(s, R)$  with a new color  $col(s)$ 
3 Choose an uncolored vertex  $d$  at distance  $R + 1$  from  $s$ 
4 Let  $P = stopBFS(d, col(s))$  and  $f$  the last vertex of  $P$ 
5 If  $P$  is of length less than  $2R + 4K + 2$  then
   |   /* Not deep enough tree: no laminar is crossed                               */
6   |   Color all vertices visited by  $stopBFS(d, col(s))$  with color  $lam$ 
7 else if  $\exists w, a$  s.t.  $col(w) \neq col(s)$  and  $h \in r3K(P)$  and  $d(w, a) = K + 1$  and
   |    $d(w, P) = K + 1$  then
   |   |   /* A hub has been detected by Lemma 9 configuration                       */
8   |   |   Add to  $A$  the first vertex  $a$  of  $r3K(P)$  satisfying the above
9 else if  $f$  is at distance less than  $2K$  of  $B(s, R)$  then
   |   |   /*  $P$  is deep, found no hub and came back near the root:
   |   |   |   problematic configuration                                           */
10  |   |   Add to  $B$  the vertex  $b$  in the middle of  $P$ 
11  |   |   Color uncolored vertices in  $B_{G \setminus \{B(d, R) \cup B(f, R)\}}(P, K)$  with color  $lam$ 
12 else if  $f$  is not adjacent to a colored vertex then
   |   |   /*  $P$  is long, found no hub and doesn't come back: dead end           */
13  |   |   Add  $f$  to  $A$ 
14 else
   |   |   /*  $P$  links  $B(s, R)$  to a vertex of a color different from  $col(s)$ .
   |   |   |   The dominated vertices correspond to a laminar.                   */
15  |   |   Color uncolored vertices in  $B_{G \setminus \{B(d, R) \cup B(f, R)\}}(P, K)$  with color  $lam$ 

```

4.5 Finding laminars

At this step we have a set of unmovable hubs including all hubs of degree 1 or at least 3, and potentially those of degree 2. Moreover, the set B of movable hub-centers indicates the places where problematic configuration occur. We have to identify the laminars and their paths, keeping in mind that some new hubs of degree 2 may be detected. Each path is found by a BFS starting at an hub center and ending at the first other hub center encountered. Then we remove from the graph the vertices from the laminar, but not the hubs. For each path P linking two hub centers h and h' , the vertices from $B(P, k) - (B(h, R) \cup B(h', R))$ are removed from the graph. Hub center h is no more used when $B(h, R)$ becomes disconnected and the whole process ends when the graph consists in disconnected hubs only.

To prevent any difficulty arising from ending a shortest path with a movable hub $B(b, R)$, we start by those hubs to run BFSs. Indeed, such hubs correspond to a configuration where the quotient of the decomposition (the one supposed by Theorem 6 and the computed one, since they have the same reduced quotient) contains a cycle. If a movable hub has been used, it means that only one hub center $a \in A$ corresponding to hub-center $h \in H$ of degree at least 3 has been found, and that all other hubs are of degree 2 on the cycle and have been missed. Starting from b , the first element of $A \cup B$ which is hit is then a , whatever the direction $B(b, R)$ was left. Thus, two BFS from b to a are run which follow the ring in

opposite direction. Either the two obtained paths K -dominate all vertices of the ring, in which case b is transferred to A and the two paths added to \mathcal{Q} ; Or there exist a vertex in the ring which is not K -dominated. This vertex is then at distance at most $K + 2r$ of some $h \in H$ (cf Appendix for a proof). It is thus added to A and b is deleted from B .

Once the movable centers have been considered, no other places with problematic configurations are left. One therefore just has to draw shortest paths between vertices of A , and Lemma 12 ensures that they cover the laminars of (H, \mathcal{P}) . The only difficulty is again that a hub of degree 2 that had not been discovered by *FindHubs* may this time be discovered by *FindLaminars* because Lemma 9 configuration is encountered. In that case, this degree 2 hub center is added to A and a new BFS is run from it. See pseudo-code of *FindLaminars* in Appendix.

► **Lemma 14.** *Suppose that (H, \mathcal{P}) has at least a hub of degree 3, and $\ell(H, \mathcal{P}) > 2R + 8K + 2r + 18k + 4$. Suppose that *FindLaminars* is run on sets A and B returned by *FindHubs*. Then it ends with every vertex deleted or marked as undeletable.*

As shown in the appendix, Lemmas 13 and 14 imply Theorem 6.

5 Embedding and distance labeling

5.1 Circle embedding with bounded distortion

Corollary 5 is a consequence of Theorem 4 and the two following propositions.

► **Proposition 15.** *Any graph G having a circle embedding with distortion γ has a shortest path or an isometric cycle with eccentricity $\lfloor \gamma/2 \rfloor$ at most.*

► **Proposition 16.** *Given a graph G and an isometric cycle with eccentricity k in G , an embedding of G in a circle with distortion $O(k \cdot cd(G))$ can be computed in polynomial time.*

Proof sketch. The construction of the embedding is similar to that of [5] with Euler tours of trees of depth k rooted in the cycle (see [5]). We then obtain an embedding of the graph in a cycle of length $2n$ at most that can be easily embedded in a circle with same length. The distortion of an edge uv of G is then at most twice the size of the union S of trees rooted on the shortest path of the cycle from the root u' of the tree of u to the root v' of the tree of v . As we have $d_G(u', v') \leq 2k + 1$, the diameter of S is at most $4k + 1$. Now consider an embedding of G in a circle C with distortion $\gamma = cd(G)$. Two nodes of S are embedded at distance at most $\gamma(4k + 1)$ in the circle and different nodes are at distance 1 at least. We thus have $|S| \leq \gamma(8k + 2)$, and our embedding has distortion $O(\gamma k)$. ◀

5.2 Distance labeling for general hub-laminar decomposition

A decomposition of a graph G in an hub-laminar allows to compute a compact representation of distances in G with additive distortion. A distance labeling is said to be c -additive and have s bit labels when the label L_u assigned to a node u contains at most s bits and for all pairs of nodes u, v , a distance estimation \widehat{d}_{uv} can be computed from L_u and L_v such that $d(u, v) \leq \widehat{d}_{uv} \leq d(u, v) + c$. Corollary 7 is a consequence of Theorem 6 and the following proposition.

► **Proposition 17.** *Given a (r, k) -hub-laminar decomposition (H, \mathcal{P}) with λ laminars of a graph G , a $\max(4k, 2r)$ -additive distance labeling with $O(\lambda \log n)$ bit labels can be computed in polynomial time.*

6 Acknowledgments

The authors thank Michel Habib for inspiring discussions about k -laminar graphs, and Eric Bapteste, Philippe Lopez and Chloé Vigliotti for raising the problem of identifying complex laminar structures in biological graphs.

References

- 1 Patrice Assouad. étude d'une dimension métrique liée à la possibilité de plongements dans \mathbf{R}^n . *C. R. Acad. Sci. Paris Sér. A-B*, 288(15):A731–A734, 1979.
- 2 Mihai Badoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *STOC 2005*, pages 225–233. ACM, 2005.
- 3 Mihai Badoiu, Kedar Dhamdhere, Anupam Gupta, Yuri Rabinovich, Harald Räcke, R. Ravi, and Anastasios Sidiropoulos. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. In *SODA 2005*, pages 119–128. SIAM, 2005.
- 4 Etienne Birmelé, Fabien de Montgolfier, and Léo Planche. Minimum eccentricity shortest path problem: An approximation algorithm and relation with the k -laminarity problem. In *Combinatorial Optimization and Applications - 10th International Conference, COCOA 2016, Hong Kong, China, December 16-18, 2016, Proceedings*, pages 216–229, 2016.
- 5 Feodor F. Dragan and Arne Leitert. On the minimum eccentricity shortest path problem. In *Algorithms and Data Structures - 14th International Symposium, WADS 2015, Victoria, BC, Canada, August 5-7, 2015. Proceedings*, pages 276–288, 2015.
- 6 Michael R. Fellows, Fedor V. Fomin, Daniel Lokshtanov, Elena Losievskaja, Frances A. Rosamond, and Saket Saurabh. Distortion is fixed parameter tractable. *TOCT*, 5(4):16:1–16:20, 2013.
- 7 Cyril Gavoille and Olivier Ly. Distance labeling in hyperbolic graphs. In *ISAAC 2005*, volume 3827 of *Lecture Notes in Computer Science*, pages 1071–1079. Springer, 2005.
- 8 Cyril Gavoille, David Peleg, Stéphane Pérennes, and Ran Raz. Distance labeling in graphs. *J. Algorithms*, 53(1):85–112, 2004.
- 9 Piotr Indyk. Algorithmic applications of low-distortion geometric embeddings. In *FOCS 2001*, pages 10–33. IEEE Computer Society, 2001.
- 10 Piotr Indyk and Jiří Matoušek. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry, Second Edition.*, pages 177–196. Chapman and Hall/CRC, 2004.
- 11 Daniel Lokshtanov. Finding the longest isometric cycle in a graph. *Discrete Applied Mathematics*, 12(157):2670–2674, 2009.
- 12 David Peleg. Proximity-preserving labeling schemes. *Journal of Graph Theory*, 33(3):167–176, 2000.
- 13 Jimmy H et al. Saw. Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Phil. Trans. R. Soc. B*, 370(1678):20140328, 2015.
- 14 Torsten Thomas, Jack Gilbert, and Folker Meyer. Metagenomics-a guide from sampling to data analysis. *Microbial informatics and experimentation*, 2(1):3, 2012.
- 15 Mikkel Thorup and Uri Zwick. Approximate distance oracles. *J. ACM*, 52(1):1–24, 2005.
- 16 Finn Völkel, Eric Bapteste, Michel Habib, Philippe Lopez, and Chloe Vigliotti. Read networks and k -laminar graphs. arXiv:1603.01179, Mar 2016.

Appendix to *Decomposing a Graph into Shortest Paths with Bounded Eccentricity*: Proofs and pseudocodes

► **Definition 18** (k -neighbor operator $_'$). . When dealing with a path P (resp. a cycle C) that k -dominates a set U , for any vertex $u \in U$, we define u' as one of the vertices of P (resp. C) such that $d(u, u') \leq k$. We say that u' is a k -neighbor of u in P (resp. C). All proofs using this operator stand whatever vertex is chosen for u' when many vertices of P (resp. C) may be chosen.

7 Isometric cycle

This section aims to prove Lemma 8.

For any cycle C and any pair of vertices a and b , we denote by $C_{a,b}$ and $C_{b,a}$ the two paths in C linking a and b .

► **Lemma 19.** *Let G be a graph with an isometric cycle k -dominating G . Let u and v be any two vertices, and u' and v' their k -neighbor as in Definition 18. Let u, v such that u (resp. v) is at distance at most k of u' (resp. v').*

Every path between u and v $2k$ -dominates either $C_{u',v'}$ or $C_{v',u'}$.

Proof. Let P be a path between u and v . Suppose that P does not $2k$ -dominate some vertex b on the path $C_{v',u'}$ and consider any vertex a in $C_{u',v'}$.

Without loss of generality, assume that u' (resp. v') is in the path $C_{a,b}$ (resp. $B = C_{b,a}$).

Then u at distance at most k of $C_{a,b}$ and v at distance at most k of $C_{b,a}$. Moreover, as every vertex of G is at distance at most k of one of those two paths, there exist c and d that are adjacent vertices in P such that c is at distance at most k of $c' \in C_{a,b}$ and d at distance at $d' \in C_{b,a}$.

As $d(c', d') \leq d(c', c) + d(c, d) + d(d, d') \leq 2k + 1$ and C is an isometric cycle, either $C_{c',d'}$ or $C_{d',c'}$ is of length at most $2k + 1$ and is thus $2k$ -dominated by $\{c, d\}$. But b and a are not in the same path on C between c' and d' , and b cannot be $2k$ -dominated by P , so that a is thus $2k$ -dominated by P .

The previous claim being true for every a in $C_{u',v'}$, the lemma is verified. ◀

► **Lemma 20.** *Let G be a graph with an isometric cycle $C = c_1, \dots, c_p$ k -covering G .*

Let D be a longest isometric cycle of G . (every other isometric cycle of G is of length at most $|D|$)

Every vertex of C is at distance at most $5k$ of D , furthermore, if D is of length more than $12k + 2$ then every vertex of C is at distance at most $2k$ of D .

Proof. Let D be a longest isometric cycle, assume that there exists c_l in C such that no vertex of D is at distance less than $2k$ of c_l .

Let c_i (resp. c_j) be vertices at distance less than k of D and such that C_{c_{i+1}, c_l} (resp. $C_{c_l, c_{j-1}}$) contains no vertex at distance less than k of D .

Let a (resp. b) the vertex of D at distance less than k of c_i (resp. c_j). By lemma 19 we know that $D_{a,b}$ and $D_{b,a}$ both $2k$ dominate either C_{c_i, c_j} or C_{c_j, c_i} . As we assumed that c_l was at distance more than $2k$ of D , we have that both $D_{a,b}$ and $D_{b,a}$ $2k$ dominate C_{c_j, c_i} .

Let m be the vertex in the middle of $D_{a,b}$ and c_m the vertex of C at distance less than k of m . By the previous hypothesis, c_m is in C_{c_j, c_i} .

78:14 Decomposing a Graph into Shortest Paths with Bounded Eccentricity

As C_{c_j, c_i} is $2k$ -dominated by $D_{b,a}$, there exist m' in $D_{b,a}$ at distance less than $2k$ of c_m and :

$$d(m, m') \leq d(m, c_m) + d(c_m, m') \leq 3k \quad (1)$$

As D is an isometric cycle, $|D_{m,m'}| \leq 3k$ or $|D_{m',m}| \leq 3k$. Assume without loss of generality that the first inequality holds and that b is in $D_{m,m'}$. Then,

$$|D_{m,b}| \leq 3k \quad (2)$$

$$|D_{a,b}| \leq 6k + 1 \quad (3)$$

By taking m in the middle of $D_{b,a}$ (instead of $D_{a,b}$), we show also that $D_{b,a}$ is of length less or equal than $6k + 1$.

Finally, D is of length at most $12k + 2$.

Let's now assume that D is of length $p \leq 12k + 2$. Consider two opposite vertices u, v in D , that is at distance at least $\lfloor \frac{p}{2} \rfloor$.

Let c_i (resp. c_j) in C at distance less than k of u (resp. v). Then,

$$d(c_i, c_j) \geq d(u, v) - d(c_i, u) - d(v, c_j) \geq \lfloor \frac{p}{2} \rfloor - 2k \quad (4)$$

It follows that,

$$|C_{c_j, c_i}| \leq |C| - d(c_i, c_j) \leq \lceil \frac{p}{2} \rceil + 2k \leq 8k + 1 \quad (5)$$

$$|C_{c_i, c_j}| \leq 8k + 1 \quad (6)$$

Hence, for every c_l in C , one of those inequalities holds :

$$d(c_i, c_l) \leq 4k \quad (7)$$

$$d(c_j, c_l) \leq 4k \quad (8)$$

As u (resp. v) is at distance k of c_i (resp. c_j), one of those inequalities holds :

$$d(u, c_l) \leq d(u, c_i) + d(c_i, c_l) \leq 5k \quad (9)$$

$$d(v, c_l) \leq 5k \quad (10)$$

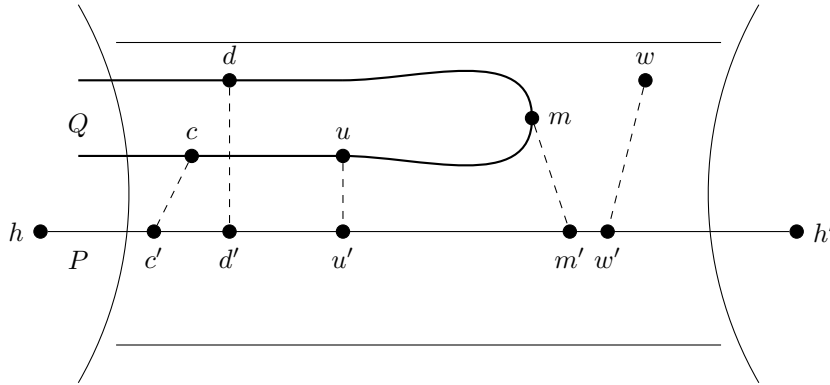
◀

Lemma 20 trivially implies Lemma 8 as C k -dominates G .

8 Proofs of the structural lemmas

This section details the proof of Lemmas 9 to 12, as well as an additional technical lemma. The proofs are given in the order corresponding to the extended abstract. Notice that Lemmas 9 and 10 rely on Lemma 12 and on the technical lemma given in the end of the section, but that the converse is not true, ensuring the validity of all of them.

► **Lemma** (Hub trigger, Lemma 9). *Consider three numbers r, k and $K \geq 3k$. If there exists*



■ **Figure 4** Illustration of the proof of Lemma 9

- a shortest path Q from a to b
 - a vertex $u \in V(Q)$ such that $d(a, u) > K + 6k$ and $d(b, u) > K + 6k$
 - a vertex v such that $d(u, v) = K$
 - a vertex w such that $vw \in E(G)$ and $d(Q, w) = K + 1$ and $d(u, w) = K + 1$,
- then any (r, k) -hub-laminar decomposition (H, \mathcal{P}) has a hub center $h \in H$ with $d(u, h) \leq K + r$.

Proof. For the sake of contradiction, suppose no such hub exists. Then u must be in a laminar with path P from h to h' and there exists $u' \in P$ such that $d(u, u') \leq k$.

Suppose first that $w \in B(P, k)$. Consider $w' \in P$ such that $d(w, w') \leq k$, and suppose w.l.o.g. that $u' \in P_{hw'}$. Suppose there exist $z \in Q$ and $z' \in P_{w'h'}$ with $d(z, z') \leq k$. Lemma 12 applied to Q_{uz} then implies that $d(w, Q) \leq 3k \leq K$, which is a contradiction. Thus, Q cannot intersect $B(P_{w'h'}, k)$.

We can thus define m' as the k -neighbor of Q on $P_{u'w'}$ furthest from u' and $m \in Q$ such that $d(m, m') \leq k$, as shown in Figure 4. As $d(w, Q) > 3k$, $d(w', m') \geq k + 1$ so that, using $d(u', w') \leq K + 2k + 1$, $d(u', m') \leq K + k$. It implies that $d(u, m) \leq K + 3k$. The hypothesis on $d(u, a)$ and $d(u, b)$ then imply that there exist two vertices c and d on Q at distance $3k + 1$ of m . As u is at distance greater than $K + r$ of any hub-center, $Q_{c,d} \subset B(P, k)$. Thus, the fact that m' is the furthest k -neighbor of Q on P contradicts Lemma 21 applied to c , m and d .

We therefore have $w \notin B(P, k)$. Consider then the last node x in $B(P, k)$ on the path from u to v . By assertion 4. of the definition of a hub-laminar decomposition, there exist a hub center h such that $x \in B(h, r)$, and thus $d(u, h) \leq K + r$.

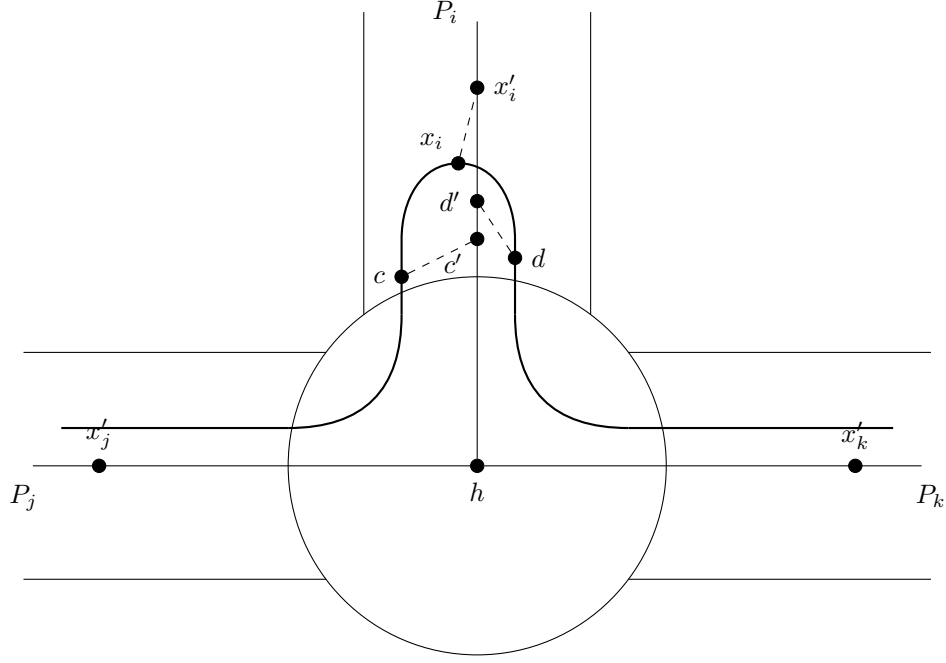
◀

► **Lemma (Degree ≥ 3 Hub Detection, Lemma 10).** *If a graph admits an (r, k) -hub-laminar decomposition (H, \mathcal{P}) , and has a hub $h \in H$ whose degree is at least 3, then for any*

- $K \geq 3k$
- shortest path Q from a to b
- vertex $u \in V(Q)$ such that $d(a, u) > r + 4K + 9k + 2$ and $d(b, u) > r + 4K + 9k + 2$ and $d(u, h) \leq K$

there exists

- a vertex $x \in V(Q)$ such that $d(a, x) > K + 6k$ and $d(b, x) > K + 6k$
- a vertex v such that $d(x, v) = K$
- a vertex w such that $vw \in E(G)$ and $d(Q, w) = K + 1$ and $d(x, w) = K + 1$,



■ **Figure 5** Illustration of the proof of Lemma 10

Proof. Consider three paths P_i, P_k, P_l of \mathcal{P} with h as an endpoint and vertices x'_i, x'_j, x'_l on those paths, each at distance $r + K + 3k + 2$ from h .

Assume first that those three vertices are at distance at most K of respectively x_i, x_j, x_l , vertices of Q . None of the last three vertices belongs to the hub $B(h, r)$ as $d(h, x_i) \geq d(h, x'_i) - d(x'_i, x_i) \geq r + 3k + 2$. Moreover, we may assume w.l.o.g that x_j, x_i and x_l are in that order in Q . There exist therefore a maximal subpath Q_{cd} of Q that is part of $B(P_i, k) \setminus B(h, r)$ and that contains x_i .

Let c' and d' be vertices of P_i such that $d(c, c') \leq k$ and $d(d, d') \leq k$. Then $d(h, c') \leq d(h, c) + k \leq r + k + 1$ and similarly for c' . As $d(h, x'_i) > r + k + 1$, Lemma 21 applies to c, x_i and e and implies that $d(c, x_i) \leq 3k$ or $d(d, x_i) \leq 3k$. In both cases, as $d(h, c) = d(h, d) = r + 1$ and $d(x_i, x'_i) \leq K$, $d(h, x'_i) \leq r + K + 3k + 1$, which is a contradiction.

One of the three vertices x'_i, x'_j or x'_k is therefore at distance more than K from Q , for instance x_i . When following P_i from h to x_i , let v be the last vertex at distance K from Q , w the following vertex of P_i and x a vertex of Q such that $d(x, v) = K$. Then $d(Q, w) = K + 1$ and, assuming w.l.o.g that $x \in Q_{u,b}$,

$$d(h, v) \leq r + K + 3k + 2 \quad (11)$$

$$d(u, x) \leq d(u, h) + d(h, v) + d(v, x) \leq r + 3K + 3k + 2 \quad (12)$$

$$d(x, b) = d(u, b) - d(u, x) \geq K + 6k \quad (13)$$

◀

► **Lemma (Hub in the dead-end, Lemma 11).** *Consider the graph G' induced by a sequence of incident hubs and laminar $H_1, L_1, H_2, \dots, H_z$, such that h_1 and h_z are at distant of at least $2R + r + 2$. Suppose moreover that H_z is a hub of degree 1 and all other hubs but H_1 are of degree 2.*

Let d in L_1 be at distance at most $R + r$ of h_1 and f a vertex of G' the furthest from d . f is then at distance at most $2r + 2k$ from h_z .

Proof. *Case 1: $z = 2$*

Consider the laminar path between h_1 and h_2 and let v_d (resp. v_f) be a vertex on that path at distance at most k of d (resp. f). Such a vertex v_f might not exist if v_f is at a distance between k and r of h_1 or h_2 . In the second case, we have the result, in the first case:

$$d(d, f) \leq d(d, h_1) + d(h_1, f) \leq R + 2 \quad (14)$$

$$d(d, h_2) \geq d(h_1, h_2) - d(d, h_1) \geq 2R + r + 2 - R - r > R + 2 \quad (15)$$

which contradicts the fact that f is the furthest vertex from d . It can be shown by a very similar argument that v_f is closer to h_2 than v_d .

Then,

$$d(d, f) \leq d(v_d, v_f) + 2k$$

But

$$d(d, f) \geq d(d, h_1) \geq d(v_d, h_1) - k$$

so that

$$d(v_d, v_f) + 3k \geq d(v_d, h_1) \quad (16)$$

$$d(f, h_1) \leq 4k \quad (17)$$

Case 2: $z > 2$

In that case,

$$d(d, f) \leq d(d, h_{z-1}) + d(h_{z-1}, v_f) + k$$

But, as any shortest path between d and h_z has to meet $B(h_{z-1}, r)$,

$$d(d, f) \geq d(d, h_z) \geq d(d, h_{z-1}) + d(h_{z-1}, h_z) - 2r$$

Thus

$$d(h_{z-1}, v_f) + k + 2r \geq d(h_{z-1}, h_z) \quad (18)$$

$$d(f, h_z) \leq 2k + 2r \quad (19)$$

◀

► **Lemma** (Path local covering, Lemma 12). *Consider a shortest path P (say, from h to h'). Let Q be a path from u to v contained in $B(P, k)$.*

Assume there exists $u' \in P$ and $v' \in P$ such that $d(u, u') \leq k$ and $d(v, v') \leq k$.

Then every vertex of $P_{u'v'}$ is at distance at most $2k$ from Q .

Furthermore, every vertex of $B(P_{u'v'}, k)$ is at distance at most $3k$ of Q .

Proof. Let us define $x_0 = u$, $x_s = v$ and $Q = x_0, \dots, x_s$.

The second assertion of the lemma is straightforward given the first one. To prove the latter, we define, for all l between 0 and s , the subpath $Q_l = x_0, x_1, \dots, x_l$ and x'_l in P such

78:18 Decomposing a Graph into Shortest Paths with Bounded Eccentricity

that $d(x_\ell, x'_\ell) \leq k$

Let us show by induction on ℓ that every vertex of P between u' and x'_ℓ , is at distance at most $2k$ of P_ℓ .

- For $\ell = 0$, $Q_0 = x_0 = u$ and $x'_0 = u'$. As u' is at distance k of u , the result is true for $\ell = 0$.

- Let l in $(1...s)$ such that the property is verified for $\ell - 1$.

Every vertex y of $P_{u'x'_{\ell-1}}$ is at distance at most $2k$ of $P_{\ell-1}$ by induction hypothesis, and thus at distance at most $2k$ of P_ℓ .

Moreover, by the triangle inequality:

$$d(x'_{\ell-1}, x'_\ell) \leq d(x'_{\ell-1}, x_{\ell-1}) + d(x_{\ell-1}, x_\ell) + d(x_\ell, x'_\ell) \leq 2k + 1 \quad (20)$$

As the sub-path of P between $x'_{\ell-1}$ and x'_ℓ is a shortest path, it follows that, for every vertex y of $P_{x'_{\ell-1}x'_\ell}$,

$$d(x'_{\ell-1}, y) \leq k \text{ or } d(x'_\ell, y) \leq k, \quad (21)$$

meaning that y is at distance at most $2k$ of $P_{\ell-1}$ or of x_ℓ .

The property is verified by induction, and the lemma follows for $\ell = s$. ◀

A last technical lemma on the behavior of shortest paths is needed for the proof of the previous lemmas and the validity of the Algorithm.

► **Lemma 21.** *Consider a shortest path Q in the graph induced by $B_G(P, k)$ with $P \in \mathcal{P}$ and three successive nodes a, m, b on Q with a', m', b' on P such that $d_G(a, a') \leq k$, $d_G(m, m') \leq k$, $d_G(b, b') \leq k$.*

If a' is between b' and m' ($a' \in P_{m', b'}$), then we have $d_G(a, m) \leq 3k$.

Proof. The following relations derive easily from the hypothesis:

$$d_G(b', m') = d_G(b', a') + d_G(a', m') \quad (22)$$

$$d_G(b, a) = d_G(b, m) + d_G(m, a) \quad (23)$$

$$d_G(b, a) \leq d_G(b', a') + 2k \quad (24)$$

$$d_G(b', m') \leq d_G(b, m) + 2k \quad (25)$$

$$d_G(m, a) \leq d_G(m', a') + 2k \quad (26)$$

It follows from equations 23 and 24 that

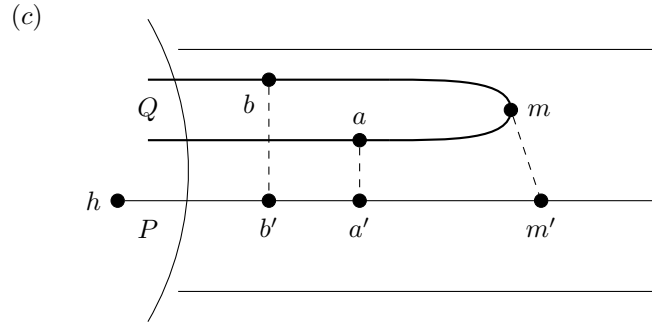
$$d_G(m, a) \leq d_G(b', a') + 2k - d_G(b, m) \quad (27)$$

Using Equation 22,

$$d_G(m, a) \leq d_G(b', m') - d_G(a', m') + 2k - d_G(b, m) \quad (28)$$

Using Equation 25,

$$d_G(m, a) \leq 4k - d_G(a', m') \quad (29)$$



■ **Figure 6** Proof of Lemma 21

Using Equation 26,

$$d_G(m, a) \leq 6k - d_G(m, a) \quad (30)$$

Finally, we get

$$d_G(m, a) \leq 3k \quad (31)$$

◀

9 Proof of the algorithm validity

9.1 Pseudocode of *FindHubs* and proof of Lemma 13

```

1 FindHubs
   Input: A graph  $G$ , integers  $R$  and  $K$ 
   Output: Two set of vertices  $A$  and  $B$ 
2 Choose any vertex  $s_0$ 
3 Run a BFS rooted in  $s_0$  and choose  $s$  as a deepest vertex
4 Run  $NextHub(G, R, K, s)$  until a vertex  $a$  is added to  $A$ 
5 Uncolor all vertices
6 Set  $A = \{a\}$  and  $B = \emptyset$ ;
7 Color every vertex in  $B(a, R)$  with color  $col(a)$ 
8 While  $\exists a \in A$  and an uncolored vertex  $d \in G$  such that  $d(a, d) = R + 1$  do
9   Apply  $NextHub(G, R, K, a)$ 
10  If a new hub  $a'$  has been discovered then
11    Add it to  $A$  or  $B$  depending on its movable status
12    Color every vertex in  $B(a', R)$  with color  $col(a')$ 

```

► **Lemma 22** (Computed hubs of degree $\neq 2$ are close to those of H). *Consider a graph G having a (r, k, ℓ, λ) hub-laminar decomposition (H, \mathcal{P}) with $\ell > 2R + 8K + 2r + 18k + 4$, and a hub center $h_0 \in H$ with hub degree three or more. Suppose that *FindHubs* is called with a starting node s such that s is at distance at most $K + r$ from a hub center h with hub degree at least 3, then:*

(i) *For every vertex a added in A at line 8, there is a hub center $h \in H$ of hub degree at least 2 at distance at most $K + r$ from a .*

(ii) *For every vertex a added in A at line 12, there is a hub center $h \in H$ of hub degree 1 at distance at most $2(k + r)$ from a .*

Proof. (i) This is a direct result of lemma 9.

(ii) Let f be added to A at line 12.

By induction, d is in a laminar L with endpoints h and h' , such that $B(h, r)$ is in $B(a, R)$. Therefore $P_{d,f}$ is such that f is in L or $P_{d,f}$ goes through h' .

Assume h' of degree at least 2. Then there exists a second laminar L' incident to h' , with path $P' = (h' = v'_0, v'_1, \dots, v'_{l'})$. Moreover,

$$d(d, f) \geq d(d, v'_{l'}) \geq d(d, h') + d(h', v'_{l'}) - 2r \geq d(d, h') + l - 2r \quad (32)$$

and,

$$d(h', f) \geq d(d, f) - d(d, h') \geq l - 2r \quad (33)$$

As any path from d to f has a vertex in $B(h', r)$, h' is not of degree more than 3, as otherwise Lemma 10 ($K \geq 3k$) would imply that h' should have been detected at line 8.

By induction we deduce that there is no hub of degree more than 3 in $G \setminus B(a, R)$ and that any path in $G \setminus B(a, R)$ starting on d goes through the same set of hubs in the same order. This set either end with an hub of degree 1 or with the hub h (the last laminar intersects $B(a, R)$).

In the second case, let v_0, v_1, \dots, v_z be the path associated with the laminar, and v_j the vertex closest to v_z and not in $B(a, R)$.

$$d(d, v_j) \geq d(d, v_0) + d(v_0, v_j) - 2r \quad (34)$$

$$d(d, f) \leq d(d, v_f) + k \leq d(d, v_0) + d(v_0, v_f) + k \quad (35)$$

$$d(d, v_0) + d(v_0, v_f) + k \geq d(d, v_0) + d(v_0, v_j) - 2r \quad (36)$$

$$d(v_0, v_f) \geq d(v_0, v_j) - 2r - k \quad (37)$$

$$d(v_f, a) \leq R + 2r + k \quad (38)$$

$$d(f, a) \leq R + 2r + 2k \quad (39)$$

In the first case, let h'' be the ending hub of degree 1 and v''_0, v''_1, \dots, h'' be the path associated with the laminar, by lemma 11, we have :

$$d(f, h'') \leq 2k + 2r \quad (40)$$

◀

The following lemma is needed to prove the converse result and implies that *FindHubs* terminates.

► **Lemma 23** (Uncolored vertices are close to A). *Consider a graph G having a (r, k, ℓ, λ) hub-laminar decomposition (H, \mathcal{P}) with $\ell > 2R + 8K + 2r + 18k + 4$ such that at least one hub has degree three or more.*

The algorithm FindHubs ends with every vertex colored or at distance $3R + 4K$ at most from a vertex a of A

Proof. Let y be an uncolored vertex at distance at least $3R + 4K + 1$ of a and x_0 the closest colored vertex. We denote by $Q = (x_0, x_1, \dots, x_t = y)$ the shortest path from x_0 to y . All vertices in Q but x_0 are then uncolored.

x_0 cannot have been colored by a vertex of A , as x_1 would have been selected at line 8 and thus colored. x_0 has therefore color lam . It cannot have been colored at Line 6 as y would then also be colored. It has therefore been colored by a path $P_{d,f}$.

x_0 is at distance K of some vertex u in $P_{d,f}$, otherwise x_1 would have been colored by u . It implies that u is at distance less than $3K$ of d or f , otherwise x_1 would have been colored at line 8. We may suppose w.l.o.g that it is at distance at most $3K$ from d , so that $d(d, x_0) \leq 4K$.

We have that d is at distance $R + 1$ of some $a \in A$. Vertices added to A at line 10 have no uncolored vertex at distance $R + 1$. Indeed, Lemmas 9 and the size of L imply that a vertex would then have been added at line 8. By Lemma 22, a is thus at distance more than $r + K$ of some hub center $h \in H$, so that $R > 2(K + r)$ implies that d is at distance at least $K + r$ from h . By definition of an hub-laminar decomposition, d is in a laminar L , which associated path is denoted by $P = (v_0 = h, v_1 \dots v_r)$.

Let v_{i_d} and v_{i_1} be vertices of P at distance less than k of d and x_1 . As

$$d(d, y) \geq d(a, y) - d(d, a) \geq 3R + 4K + 1 - R - 1 \geq 2R + 4K, \quad (41)$$

$$d(d, f) \geq \min(d(d, y), l - 2R + 2) \geq 2R + 4K \quad (42)$$

Lemma 12 thus implies that every vertex at distance k of the path $v_{i_d}, \dots, v_{i_d+2R+4K-2k}$ is K covered by $P_{d,f}$. So i_1 is either lower than i_d or greater than $i_d + 2R + 4K - 2k$.

But

$$d(v_{i_d}, v_{i_1}) \leq d(v_{i_d}, d) + d(d, x_1) + d(x_1, v_{i_1}) \leq 4K + 2k + 1 \quad (43)$$

$$d(h, v_{i_1}) \leq d(h, v_{i_d}) + 4K + 2k + 1 \leq d(h, v_{i_d}) + 2R + 4K - 2k \quad (44)$$

Thus, $i_1 \leq i_d$.

As $B(v_{i_d+2R}, K)$ disconnects L and is entirely colored, Q is entirely in L . Let us denote v_{i_j} a vertex of P at distance k of x_j for every $2 \leq j \leq t$. Then for every $1 \leq j < t$, $d(v_{i_j}, v_{i_{j+1}}) \leq 2k + 1$, so that all indices i_j are lower than i or all are greater than $i + 2R + 4K - 2k$. It follows that they are all lower than i_d .

But

$$i_d = d(h, v_i) \leq d(h, a) + d(a, d) + d(d, v_{i_d}) \leq 2R + 2k - r$$

so that $d(h, v_{i_t}) \leq 2R + 2k - r$.

Finally,

$$d(a, y) \leq d(a, h) + d(h, v_{i_t}) + d(v_{i_t}, y) \leq R - r + 2R + 2k - r + k < 3R + 4K + 1$$

which is a contradiction. ◀

► **Lemma 24** (Hubs of H of degree $\neq 2$ are close to computed ones). *Consider a graph G having a (r, k, ℓ, λ) hub-laminar decomposition (H, \mathcal{P}) with $2R + 7K < \ell - 2r - 3k - 3$ such that at least one hub of degree three or more.*

When FindHubs terminates:

(i) *For every hub $h \in H$ of degree at least 3, there exists a vertex a in A at distance at most $K + r$ from h .*

(ii) *For every hub $h \in H$ of degree 1, there exists a vertex a in A at distance at most $2(k + r)$ from h .*

Proof. (i) Assume that h is at distance more than $\max(R + r + 5K + 9k + 3, 3R + 4K)$ of every vertex a of A . By Lemma 23, h is of color *lam*. Therefore there is a path $P_{d,f}$ and vertices a, a' in A , u in $P_{d,f}$ such that :

$$d(d, a) \leq R + 1 \quad (45)$$

$$d(f, a') \leq R + 1 \quad (46)$$

$$d(h, a) \geq R + r + 5K + 9k + 3 \quad (47)$$

$$d(h, a') \geq R + r + 5K + 9k + 3 \quad (48)$$

$$d(u, h) \leq K \quad (49)$$

By combining those inequalities :

$$d(d, u) \geq d(h, a) - d(d, a) - d(u, h) \geq r + 4K + 9k + 2 \quad (50)$$

$$d(f, u) \geq d(h, a') - d(d, a') - d(u, h) \geq r + 4K + 9k + 2 \quad (51)$$

By Lemma 10, there is $vw \in E(G)$ and x at distance more than $K + 6k$ from d and f such that $d(x, v) = K$, $d(P, w) = K + 1$. It results that h should have been detected at line 14 and colored by a vertex of A .

Assume that h is at distance less than $R + r + 5K + 9k + 3$ of a vertex a of A added at line 12. By lemma 22, there exist a hub of degree 1 of center h' such that $d(a, h') \leq K + r$. Therefore, $d(h, h') \leq d(h, a) + d(a, h') \leq R + 2r + 6K + 9k + 3$, which is impossible.

Let h be at distance less than $R + r + 5K + 9k + 3$ of a added at line 8 of *NextHub*. a is at distance at most $K + r$ of a hub h' of degree 3 or more. We then have :

$$d(h, h') \leq R + 2r + 6K + 9k + 3 < l \quad (52)$$

Thus $h = h'$, so that h is at distance at most $K + r$ of a .

ii) Let h be an hub of degree 1. Suppose first that h is at distance less than $3R + 4K$ of a vertex a of A . By a proof similar to (i), a was added at line 12 of *NextHub* and is at distance at most $K + r$ of h .

Assume now that h is at distance more than $3R + 4K$ of every vertex a of A . Then there is a path $P_{d,f}$, and u in $P_{d,f}$ at distance at most K of h . We have,

$$d(d, u) \geq d(a, h) - d(a, d) - d(u, h) > 3k \quad (53)$$

$$d(f, u) > 3k \quad (54)$$

Let y and y' be the vertices of $P_{d,f}$ at distance $3k + 1$ of u . Let L be the laminar having h as an endpoint and P the path associated to L .

As $3k + 1 < l$, y and y' are in L . Let v (resp. v') a vertex of P_L at distance at most k of y (resp. y'). Assume w.l.o.g that v is closest to h than v' .

Then $v \in P_{h,v'}$, which leads to a contradiction by lemma 21. Thus, h cannot be at distance more than $3R + 4K$ of every vertex a of A . ◀

► **Lemma (Lemma 13).** *Suppose that (H, \mathcal{P}) has at least a hub of degree 3, and $\ell(H, \mathcal{P}) > 2R + 7K + 2r + 3k + 3$. Then, for every vertex in $a \in A$, there exist a vertex in $h \in H$ such that their distance is at most $K + 2r$. Conversely, for every $h \in H$ of degree different from 2, such a vertex a is selected in A .*

Proof. This is a direct result of lemma 22 and 24. ◀

9.2 Pseudocode of *FindLaminars* and proof of Lemma 14

```

1 FindLaminars
   Input: A graph  $G$ , integers  $R$  and  $K$ 
   Output: a hub-laminar decomposition  $(A, \mathcal{Q})$ 
2  $(A, B) = \text{FindHubs}(G, R, K)$ 
3  $\mathcal{Q} = \emptyset$ 
4 Mark all vertices as deletable
5 For each vertex  $a$  of  $A$  do
6   | Mark the vertices in  $B(a, R)$  as undeletable
7 For each  $b \in B$  do
8   | Run a BFS starting at  $b$  and stopping on the first vertex  $a \in A$ 
9   | Let  $Q_1$  be the path from  $b$  to  $a$  computed by this BFS
10  | Run a BFS starting at  $b$ , not using vertices of  $B(Q_1, K) \setminus (B(b, R) \cup B(a, R))$ 
    | and stopping in  $a$ 
11  | Let  $Q_2$  be the path from  $b$  to  $a$  computed by this BFS
12  | Compute  $G'$ , the union of the connected component of  $G \setminus B(a, R)$  containing  $b$ 
    | and  $B(a, R)$ 
13  | Color in  $G'$  the vertices of  $B(a, R)$ ,  $B(b, R)$ ,  $B(Q_1, K)$  and  $B(Q_2, K)$ 
14  | If  $\exists$  an uncolored vertex  $a$  in  $H$  then
15  |   | Add  $a$  to  $A$  Mark the vertices in  $B(a, R)$  as undeletable
16  | else
17  |   | Add  $b$  to  $A$ 
18  |   | Mark the vertices in  $B(b, R)$  as undeletable
19  |   | Delete from  $G$  the deletable vertices from  $B(Q_1, K) \cup B(Q_2, K)$ 
20  |   | Add  $Q_1$  and  $Q_2$  to  $\mathcal{Q}$ 
21 While there exists  $a \in A$  such that  $B(a, R+1) \neq B(a, R)$  do
22  | Run a BFS starting at  $a$  and stopping on the first vertex  $a' \in A$ ,  $a' \neq a$ 
23  | Let  $Q$  be the path from  $a$  to  $a'$  computed by this BFS
24  | If  $\exists w, h$  s.t.  $h \in r3K(Q)$ ,  $d(w, h) = K+1$  and  $d(w, Q) = K+1$  then
25  |   | Add to  $A$  the first vertex  $h$  of  $Q$  satisfying the above
26  |   | Mark the vertices in  $B(h, R)$  as undeletable
27  | else
28  |   | Add to  $\mathcal{Q}$  the path  $Q$  from  $a$  to  $a'$  computed by this BFS
29  |   | Delete from  $G$  the deletable vertices from  $B(P, K)$ 

```

► **Lemma (Lemma 14).** *Suppose that (H, \mathcal{P}) has at least a hub of degree 3, and $\ell(H, \mathcal{P}) > 2R + 7K + 2r + 3k + 3$. Suppose that *FindLaminars* is run on sets A and B returned by *FindHubs*. Then it ends with every vertex deleted or marked as undeletable.*

Proof. Consider the underlying decomposition (H, \mathcal{P}) of G . Let $X = \{x_1, \dots, x_\lambda\}$ be the set of midpoints of the laminar paths (for paths of odd length, an arbitrary choice is made between the two possible vertices). In the following of the proof, a laminar will be denoted by $L(x)$, where x is the unique element of X belonging to the laminar.

A $h \in H$ is said to be *covered by* A if there exist a vertex of A at distance at most $K + 2r$ of h .

Claim 1: The vertices belonging to a hub $B(h, r)$, h being covered in the end of the

algorithm, are marked as undeletable.

h being covered, some $a \in A$ is at distance at most $K + 2r$ of h . As $K + 2r + r < R$, $B(h, r) \subset B(a, R)$ and its vertices are undeletable.

Claim 2: Consider a laminar $L(x)$ which laminar path joins h_1 to h_2 . Suppose that at some point, h_1 and h_2 are covered and x is not deleted. Then all vertices of $L(x)$ are deleted or marked as undeletable when the algorithm terminates.

By the same reasoning as in Claim 1, as $(K + 2r) + (r + 3k + K) \leq R$, all vertices of $L(x)$ belonging to $B(h_i, r + 3k + K)$ are undeletable. Let $L'(x) = L(x) \setminus (B(h_1, r) \cup B(h_2, r))$ denote the central part of the laminar, that is the vertices belonging to the laminar but not to the incident hubs.

Let Q be any path added to \mathcal{Q} by the algorithm. To prove the claim, we shall prove that Q either delete all or none of the deletable vertices of $L(x)$. As $L(x)$ is connected, the While loop will thus be run until some path Q is selected which deletes them all.

Case 1: Q does not hit $L'(x)$

The deletable vertices of $L(x)$ being at distance at least $r + 3k + K$ from h_1 and h_2 , none of them is deleted.

Case 2: Q enters and exits $L'(x)$ by the same hub, say $B(h_1, r)$

Let x, y and z be vertices appearing on Q in this order, such that x and z are at distance r from h_1 and y is the vertex on Q the furthest of h_1 . Let x', y' and z' be vertices on the laminar path k -covering them. If y' is closer to h_1 than x' or z' , say x' ,

$$d(h_1, y) \leq d(h_1, x') + d(y', y) \leq d(h_1, x) + d(x, x') + d(y', y) \leq r + 2k$$

If not, Lemma 21 implies that $d(x, y) \leq 3k$. In any case, $d(h_1, y) \leq r + 3k$. Thus any vertex K -covered by Q is at distance at most $r + 3k + K$ from h_1 , that is is undeletable. None of the deletable vertices of $L(x)$ is thus deleted.

Case 3: Q enters $L'(x)$ by one hub and exits it by the other one

Let u and v be vertices on Q at distance r from respectively h_1 and h_2 . Let u' and v' on the laminar path of $L(x)$ that k -cover u and v . As $d(h_1, u') \leq r + k$ and $d(h_2, v') \leq r + k$, all deletable vertices of $L(x)$ are k -covered by the laminar subpath between u' and v' . Lemma 12 then implies that they are K -covered by Q . All of them are thus deleted.

Case 4: Q has an endpoint in $L'(x)$

Let a be that endpoint and w.l.o.g., let suppose it covers h_1 . Denote by a' a vertex on the laminar path that k -covers a . Then all vertices of $L(x)$ that are k -covered by the laminar subpath between h_1 and a' are at distance less than R of a and are thus undeletable. Concerning those which are covered by the laminar subpath between a' and h_2 , the proofs of Cases 1 to 3 can be rewritten by replacing h_1 by a' . Either Q exits $L'(x)$ by $B(h_2, r)$, covering all deletable vertices, or it exits $L(x)$ by $B(h_1, r)$, covering none of them.

Claim 3: Consider a sequence of hubs and laminars $H_1, L(x_1), H_2, \dots, L(x_{p-1}), H_p$ corresponding to a path in the quotient graph and such that the H_i 's are of degree 2 for $2 \leq i \leq p - 1$. Suppose that at some point the hub-centers of H_1 and H_p are covered, those of the others hub aren't, and x_1 is not deleted. Then all vertices of the hubs and laminars of the sequence are deleted or marked as undeletable when the algorithm terminates.

This claim is proven exactly in the same way that the preceeding one, except that in the case where the path Q crosses the whole sequence of laminars, Lemma 12 doesn't ensure that all vertices are K -covered. However, if this should not be the case, the condition of Line 24 is fulfilled, in which case a vertex covering a hub of degree 2 is added to A . The

sequence is then cut into two subsequences, on which the claim can be recursively applied. Claim 2 ensures that only a finite number of recursions are needed.

Claim 4: Consider a sequence of hubs and laminars $H_1, L(x_1), H_2, \dots, L(x_{p-1}), H_1$ corresponding to a cycle in the quotient graph and such that the H_i 's are of degree 2 for $2 \leq i \leq p-1$. Then all vertices of the hubs and laminars of the sequence are deleted or marked as undeletable when the algorithm terminates.

This situation corresponds to the problematic case, H_1 being the only hub of degree at least three in the sequence, and is thus covered by Lemma 24. If one of the hubs of degree 2 has been covered during the *FindHubs* procedure, the result is true by applying Claims 2 or 3 to the resulting subsequences.

Suppose therefore that no hub of degree 2 is covered. Then a movable b has been introduced by *FindHubs* and two paths Q_1 and Q_2 linking b to the vertex a covering the hub-center of H_1 are drawn. Note that Lemma 12 implies that Q_1 K -covers the central part of the laminars it crosses so that Q_2 has to go the other way around: if Q_1 reaches a by $L(x_1)$, Q_2 reaches it by $L(x_{p-1})$.

Consider the graph G' as introduced in the algorithm and which correspond to the union of $B(a, R)$ with the graph induced by the considered sequence of laminars and hubs. The R -neighborhoods of a and b as well as the K -neighborhood of $Q_1 \cup Q_2$ are colored.

If there is no uncoloured vertex in G' , all of them are either deleted or marked as undeletable at lines 18 and 19. Supposetherefore that there exist an uncolored vertex a . We shall prove that a covers one of the hubs of degree two, so that the fact that it is added to A at line again creates two subsequences for which the result is true by Claims 2 or 3.

If a belongs to one of the hubs of the sequence, the result is obvious, so let's suppose it belongs to some laminar $L(x_i)$.

Case 1: neither a or b belongs to $L(x_i)$

Then one of the two paths Q_1 or Q_2 has to cross $L(x_i)$, that is contains a vertex u at distance r from h_i and a vertex v at distance r from h_{i+1} . Let u' and v' be vertices of the laminar path of $L(x_i)$ k -dominating them. Then $d(h_i, u') \leq k + r$ and $d(h_{i+1}, v') \leq k + r$.

By Lemma 12, a cannot be k -covered by the laminar subpath between u' and v' as it would have been colored. It is therefore at distance k of a vertex being on the subpath between h_1 and u' or between h_2 and v' . Thus, its distance to h_i or h_{i+1} is bounded by $2k+r$, so it covers it.

Case 2: a or b belongs to $L(x_i)$

The same reasoning as in Case 1 by replacing $B(h_i, r)$ and/or $B(h_{i+1}, r)$ by $B(a, K)$ and/or $B(b, K)$. The only difference is that v cannot be close to one of those balls as $K + 2k < R$ and that v would thus have been colored.

◀

9.3 Proof of Theorem 6

Let us first show that the output of the algorithm fulfills the definition of n hub-laminar decomposition.

- Axiom 1. Each laminar links two hubs centers. The endpoints h, h' of any $P \in \mathcal{P}$ belong to H and for every other hub $h'' \in H \setminus \{h, h'\}$, $B(P, k) \cap B(h'', r+1) = \emptyset$:

The set of paths \mathcal{Q} is defined in the function *FindLaminars*. They are by definition paths between two vertices of A , the set of hubs returned by the previous function.

Let Q be a path of \mathcal{Q} with end points $a_1, a_2 \in A$. Assume the existence of a third vertex $a_3 \in A$ such that $B(a_3, R)$ intersects $B(Q, K)$, i.e. a_3 is at distance $K + R$ at most from some vertex $q \in Q$. We then have:

$$\begin{aligned} d(a_1, a_3) &\leq d(a_1, q) + K + r \\ d(q, a_2) &\geq d(a_2, a_3) - d(q, a_3) \geq L - K - r \\ d(a_1, a_2) &= d(a_1, q) + d(q, a_2) \geq d(a_1, q) + L - K - r > d(a_1, a_3) \end{aligned}$$

Without loss of generality, assume that Q started on a_1 , Q is then the shortest path between a_1 and any vertex of A in the remaining graph when computing Q . As a_3 is not visited before a_2 , some vertices between q and a_3 must have been deleted during a previous step of the algorithm. This means that Q meets an other lamina outside a hub, in contradiction with Axiom 4, which is verified as detailed below.

- Axiom 2. The laminars and the hubs cover G : $V(G) \subseteq \bigcup_{h \in H} B(h, r) \cup \bigcup_{P \in \mathcal{P}} B(P, k)$:

This is a direct result of Lemma 14.

- Axiom 3. Each lamina path is locally a shortest path. Any path $P \in \mathcal{P}$ with endpoints h and h' is a shortest path of the graph $G[B(P, k) \cup B(h, r) \cup B(h', r)]$:

Each path $Q \in \mathcal{Q}$ is locally a shortest path, each path Q with endpoint a, a' is a shortest path of the remaining graph when computing Q which contains the dumbbell $B(Q, K) \cup B(a, R) \cup B(a', R)$.

- Axiom 4. Laminars meet at hubs only. For all $i \neq j$ and $uv \in E(G)$ such that $u \in B(P_i, k)$ and $v \in B(P_j, k)$, there is a hub center $h \in H$ such that P_i and P_j both have h as endpoint and $u, v \in B(h, r)$:

It is a consequence of the proof Lemma 14, and more precisely of the arguments developed in Claim 2. Indeed, they imply that for a path Q added between a_1 and a_2 , every connected component of $G \setminus \bigcup_{a \in A} B(a, R)$ not included in $B(Q, K)$ is not hit by $B(Q, K)$. Consequently, $B(Q_1, K)$ and $B(Q_2, K)$ cannot intersect vertices that are not undeletable, that is in the hubs.

The $K + 2r$ -equivalence is a consequence of Lemma 22 and Lemma 24, which allow to build the bijection ϕ between hub centers with hub degree different from 2. Notice moreover that the decomposition (A, \mathcal{Q}) has λ hubs at most since it has no more degree 2 hubs than (H, \mathcal{P}) . Our algorithm indeed adds degree 2 hubs in two cases only. First, when the conditions of Lemma 9 are met, and the added hub is then associated by Lemma 22 to a hub of H . Second, when we encounter a self loop in the reduced quotient, i.e. a sequence of (at least 2) laminars of (H, \mathcal{P}) connected by (at least 1) hubs of degree 2, the algorithm then adds only one hub (at Line 10 according to case (c)).

Regarding the time complexity, apart from case (a), each iteration of the *while* loop in *FindHubs* corresponds to finding a hub or a lamina. There are thus $O(|A| + |\mathcal{Q}|)$ such iterations, and their overall cost is $O(\min(\lambda, n)m)$. In the iterations corresponding to Case (a), all vertices visited by StopBFS are colored: the overall cost of such iterations is thus $O(m)$. Similarly, *FindLaminars* consists in λ iterations costing $O(m)$ each.

10 Embedding and distance labeling

► **Proposition** (Proposition 15). *Any graph G having a circle embedding with distortion γ has a shortest path or an isometric cycle with eccentricity $\lfloor \gamma/2 \rfloor$ at most.*

Proof sketch. Consider an embedding of G in a circle C with distortion γ . Suppose that any shortest path of G has eccentricity greater than $\lfloor \gamma/2 \rfloor$. We first show that G contains a simple cycle that $\lfloor \gamma/2 \rfloor$ covers the graph. Given a path P , two consecutive nodes u, v of P are at distance at most γ in the circle embedding, and P thus $\lfloor \gamma/2 \rfloor$ covers any node embedded between u and v in the circle. We define the arc P_C of P in C as the smallest arc of C where nodes of P are embedded. Note that all nodes embedded in P_C are $\lfloor \gamma/2 \rfloor$ covered by P . Consider a shortest path P with longest arc P_C and let a, b denote the extremities of P_C . If P does not $\lfloor \gamma/2 \rfloor$ cover G , consider a node c at distance greater than $\lfloor \gamma/2 \rfloor$ from P . c cannot be embedded in P_C . Consider a shortest path Q from c to a in G . The choice of P implies that Q_C cannot contain P_C , it thus covers nodes embedded in the arc C_{ca} of C from c to a that avoids the interior of P_C . Similarly, the shortest path R from c to b covers nodes embedded in the arc C_{cb} of C from c to b that avoids the interior of P_C . Let a' be the first node of Q in P . Let Q' be the sub-path of Q from c to a' and let P' be the sub-path of P from a' to b . Note that the arc of $Q' \cup P'$ contains the arc in C from c to b in $Q_C \cup P_C$. Similarly, let b' be the first node of R in P . Then define R' as the sub-path of R from c to b' and Q'' as the sub-path of $Q' \cup P'$ from c to b' . Note that R'_C contains the arc from c to b in $R_C \cup P_C$. The union $Q'' \cup R'$ defines a simple cycle that $\lfloor \gamma/2 \rfloor$ covers G as $Q''_C \cup R'_C = C$.

Now consider a simple cycle S of G that $\lfloor \gamma/2 \rfloor$ covers G and has minimum length. S must be isometric: otherwise there would be a path P from a to b in S that is shorter than both paths Q and R of S from a to b . Consider the arc A of C from a to b included in P_C . Without loss of generality, Q covers the nodes embedded in the other part $C \setminus A$ of the cycle. We can then construct from $P \cup Q$ (similarly as above) a simple cycle that $\lfloor \gamma/2 \rfloor$ covers G in contradiction with the choice of S as $|P| + |Q| < |S|$. ◀

► **Proposition** (Proposition 17). *Given a (r, k) -hub-laminar decomposition (H, \mathcal{P}) with λ laminars of a graph G , a $\max(4k, 2r)$ -additive distance labeling with $O(\lambda \log n)$ bit labels can be computed in polynomial time.*

Proof. We assume that hub centers are numbered from 1 to q , $q \leq 2\lambda$. For every $u \in V(G)$, we define a hub label H_u consisting in all pairs $(h, d(u, h))$ for $h \in H$. For a node u in a hub, i.e. when there exists $h \in H$ such that $u \in B(h, r)$, we define its label L_u as its hub label, i.e. $L_u := H_u$. For a node u in a laminar, i.e. there exists $P \in \mathcal{P}$ with endpoints $h_1 < h_2$ such that $u \in B(P, k) \setminus B(\{h_1, h_2\}, r)$, we additionally store $(h_1, h_2, d_P(h_1, u'), d(u', u))$ for some $u' \in B(u, k) \cap P$ and set $L_u := (h_1, h_2, d_P(h_1, u'), d(u', u)), H_u$ (we let d_P denote the distance in the graph induced by P).

The distance $d(u, v)$ between two nodes $u, v \in V(G)$ is then estimated from their labels L_u and L_v as follows. We first compute the estimate through hub centers $g(u, v) = \min_{h \in H} d(u, h) + d(v, h)$. If L_u and L_v both begin with quadruples $(h_1, h_2, d(h_1, u'), d(u', u))$ and $(h'_1, h'_2, d(h'_1, v'), d(v', v))$ respectively with $h_1 = h'_1$ and $h_2 = h'_2$, we detect that u and v belong to the same laminar and return the distance estimate $f(u, v) = \min(g(u, v), g'(u, v))$ where $g'(u, v) = d(u', u) + |d_P(h_1, u') - d_P(h_1, v')| + d(v', v)$. Otherwise, we simply return $f(u, v) = g(u, v)$ as distance estimate.

We now prove that we always have $d(u, v) \leq f(u, v) \leq d(u, v) + \max(4k, 2r)$. By triangle inequality, we have $d(u, v) \leq d(u, h) + d(v, h)$ for all $h \in H$ and thus obtain $d(u, v) \leq g(u, v)$. In the case where u and v both belong to the same laminar $B(P, k)$, note that $g'(u, v)$ is the length of a path through vertices $u', v' \in P$ from u to v , implying $g'(u, v) \leq d(u, v)$. We thus have $d(u, v) \leq f(u, v)$ in any case. Now consider a shortest path Q from u to v . First assume Q intersects a hub: there exists $h \in H$ such that $Q \cap B(h, r) \neq \emptyset$. Consider $x \in Q \cap B(h, r)$. We

then have $d(u, v) = d(u, x) + d(x, v) \leq d(u, h) + d(h, x) + d(v, h) + d(h, x) \leq d(u, h) + d(v, h) + 2r$ implying $g(u, v) \leq d(u, v) + 2r$. Second, suppose that Q does not intersect any hub, it must then be included in a laminar according to Items 2 and 4. Consider $P \in \mathcal{P}$ with endpoints $h_1 < h_2$ such that $Q \subseteq B(P, k) \setminus B(\{h_1, h_2\}, r)$. Then u and v both belong to the laminar and their labels contain quadruples $(h_1, h_2, d(h_1, u'), d(u', u))$ and $(h_1, h_2, d(h_1, v'), d(v', v))$ respectively. Consider the sub-graph G_P induced by $B(P, k)$. By triangle inequality, we have $d_{G_P}(u', v') \leq d_{G_P}(u, u') + d_{G_P}(u, v) + d_{G_P}(v, v')$. As Q is included in G_P we have $d(u, v) = d_{G_P}(u, v)$ and we obtain $|d_P(h_1, u') - d_P(h_1, v')| = d_{G_P}(u', v') \leq d(u, v) + 2k$ and thus get $f(u, v) \leq g'(u, v) \leq d(u, v) + 4k$. In any case we have $f(u, v) \leq d(u, v) + \max(4k, 2r)$. ◀